

Pre-treatment analysis of non-rigid variations can assist robust intensity-modulated proton therapy plan selection for head and neck patients

Ying Zhang¹ | Jailan Alshaikhi² | Richard A. Amos¹ | Wenyong Tan³ |
Virginia Marin Anaya⁴ | Yaru Pang¹ | Gary Royle¹ | Esther Bär^{1,4}

¹Department of Medical Physics and Biomedical Engineering, University College London, Gower Street, London, UK

²Saudi Proton Therapy Center, King Fahad Medical City, Riyadh, Saudi Arabia

³Department of Oncology, Shenzhen Hospital of Southern Medical University Shenzhen, Guangdong, China

⁴University College London Hospitals NHS Foundation Trust, Radiotherapy Physics, London UK

Correspondence

Ying Zhang, Department of Medical Physics and Biomedical Engineering, University College London, Gower Street, London WC1E 6BT, UK.

Email: ying.zhang.18@ucl.ac.uk

Funding information

China Scholarship Council, Grant/Award Number: 201809150003; Cancer Research UK, Grant/Award Number: C7893/A28990; National Natural Science Funding of China, Grant/Award Number: 81974462

Abstract

Purpose: To incorporate small non-rigid variations of head and neck patients into the robust evaluation of intensity-modulated proton therapy (IMPT) for the selection of robust treatment plans.

Methods: A cohort of 20 nasopharynx cancer patients with weekly kilovoltage CT (kVCT) and 15 oropharynx cancer patients with weekly cone-beam CT (CBCT) were retrospectively included. Anatomical variations between week 0/week 1 of treatment were acquired using deformable image registration (DIR) for all 35 patients and then applied to the planning CT of four patients who have kVCT scanned each week to simulate potential small non-rigid variations (sNRVs). The robust evaluations were conducted on IMPT plans with: (1) different number of beam fields from 3-field to 5-field; (2) different beam angles. The robust evaluation before treatment, including the sNRVs and setup uncertainty, referred to as sNRV+R evaluation was compared with the conventional evaluation (without sNRVs) in terms of robustness consistency with the gold standard evaluation based on weekly CT.

Results: Among four patients (490 scenarios), we observed a maximum difference in the sNRV+R evaluation to the nominal dose of: 9.37% dose degradation on D_{95} of clinical target volumes (CTVs), increase in mean dose (D_{mean}) of parotid 11.87 Gy, increase in max dose (D_{max}) of brainstem 20.82 Gy. In contrast, in conventional evaluation, we observed a maximum difference to the nominal dose of: 7.58% dose degradation on D_{95} of the CTVs, increase in parotid D_{mean} by 4.88 Gy, increase in brainstem D_{max} by 13.5 Gy. In the measurement of the robustness ranking consistency with the gold standard evaluation, the sNRV+R evaluation was better or equal to the conventional evaluation in 77% of cases, particularly, better on spinal cord, parotid glands, and low-risk CTV.

Conclusion: This study demonstrated the additional dose discrepancy that sNRVs can make. The inclusion of sNRVs can be beneficial to robust evaluation, providing information on clinical uncertainties additional to the conventional rigid isocenter shift.

KEYWORDS

anatomical uncertainty, head and neck cancer, intensity-modulated proton therapy, robust evaluation

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

1 | INTRODUCTION

Intensity-modulated proton therapy (IMPT) offers the potential to limit dose to normal tissues for head and neck (H&N) cancer patients. However, anatomical variations in the radiation area increase dosimetric uncertainty during treatment delivery.^{7,8} Progressive changes due to weight loss, tumour shrinkage, and parotid glands shrinkage were reported as 3.9%–25.5% (weight loss), 20%–60% (tumor shrinkage), and 21.3%–42% (parotid glands shrinkage).^{9–11} Neck folds, neck tilts, spine flexions, and jaw and shoulder position changes also commonly occur during the H&N cancer treatment.^{12,13} These small non-rigid variations (sNRVs) cannot be simplified as rigid translations and, unlike progressive changes that are patient-specific, sNRVs occur randomly.

Current research in H&N proton therapy focuses on the development of adaptive strategies to mitigate the influence of progressive anatomical changes. In clinical practice, offline adaptive planning strategies are applied when a threshold of dose to a critical structure is reached.^{14,15} This method is effective, but delays in implementing adaptive re-plans exist due to time required for imaging, re-planning, plan approval, and plan verification. This reactive approach to adaptive therapy poses workflow challenges for the busy clinical practice. To mitigate time delay during the offline adaptive process, the use of anatomical modeling was suggested.^{16,17} Anatomical models can accurately predict the patients' progressive changes and can therefore be used to create adaptive plans in advance, which can be applied as soon as the adaption threshold is reached. Online adaption is intended for same-day application. However, due to computational limitations, online adaption either compromises on accuracy or constrains the optimizer. Matter et al.¹⁸ used an analytical pencil beam algorithm to generate plans in 10 s. However, analytical calculations overestimate the target by 10% and underestimate some organs at risk (OARs) by up to 10 Gy.¹⁹ Bobić et al.²⁰ constrained the optimizer to adjust the beamlet positions, energies, and beamlet weights to produce adapted plans. They reported a median adjustment time of 12 min excluding the time taken for deformable image registration (DIR). Lalonde et al.²¹ only adjusted the weights of the beamlets to produce adapted plans, their median adjustment time was also 12 min but included the time for DIR. When plans are adapted either online or offline, the patient position may be different from the position in the image. sNRVs not captured during imaging will still be present.

In addition to adaptive planning strategies that mitigate the dosimetric impact of anatomical variability, evaluation of plan robustness is also used.^{22,23} Set up and range uncertainty are considered in conventional robust evaluation. Treatment plan evaluation including

inter-fractional anatomical variations often uses images acquired during the treatment,^{24–26} and as such, they can only inform the planning process for a portion of the treatment delivery. A more complete robust evaluation including the possible sNRVs before treatment is crucial to design a plan that is robust toward these anatomical changes. Because sNRVs are not patient-specific, they can be included into robust evaluation to provide additional information before treatment. To our knowledge, studies have yet to reveal the dosimetric impact of sNRVs on proton therapy plans.

Range uncertainty in robust analysis evaluates the dosimetric impact of the systematic uncertainty in calculated range based on CT calibration and conversion to relative stopping power (RSP), while setup uncertainty reflects random errors throughout a course of therapy. This study focused on random errors. We aim to (1) establish the additional impact of sNRV, as a component of random setup error, over and above the rigid translation by building a distribution of possible sNRVs based on population data; (2) provide a robust evaluation method based on the probability distribution. The benefit of this new evaluation method was compared to the conventional robust evaluation, with gold-standard evaluation (after-treatment evaluation that used weekly repeated CTs) as the reference for quantification.

2 | METHODS

2.1 | Patient data

Twenty nasopharynx cancer patients with weekly repeat CT and 15 oropharynx cancer patients with weekly cone-beam CT (CBCT) who received photon therapy were recruited retrospectively. We obtained deformations between week 0 (planning CT) and week 1 of treatment (the time between planning CT and treatment week 1 is 14 days, which is the standard time for treatment planning) for all 35 patients, creating a distribution of possible sNRVs based on the method described in Section 2.2. Examples of sNRVs are shown in Appendix A. Four nasopharynx patients who have weekly repeat CTs were randomly selected as test dataset, where we applied the 35 sNRVs to their planning CT.

We evaluated the robustness of IMPT plans toward the uncertainty (see Section 2.3) applied to the test patients. We evaluated the following scenarios: (1) different number of fields from 3-field to 5-field plan; (2) different beam angles. The different beam arrangements used in this paper are listed in the upper part of Table 1 and illustrated in Appendix B. The targets (both tumor and nodal area) were split for different fields in these IMPT plans. All plans were robustly optimized using ± 3 mm setup and $\pm 3.5\%$ range uncertainty in

TABLE 1 Plan beam arrangements and dosimetric goals used in this paper

Plan beam arrangements			
Beam arrangements	Angle		
$3B_{45}$	45 180 315		
$3B_{60}$	60 180 300		
$4B_{110}$	60 110 250 300		
$4B_{120}$	60 120 240 300		
$5B$	60 110 180 250 300		
Dosimetric goals of the treatment plans			
Structure	Metric goal under uncertainty		Robust optimized
High-risk CTV	D_{95}	> 95% of prescription dose (72.6Gy)	Yes
Low-risk CTV	D_{95}	> 95% of prescription dose (63Gy)	Yes
CTV	D_2	< 107% of prescription dose	Yes
Spinal cord	D_{max}	<45 Gy	Yes
Brainstem	D_{max}	<55 Gy	Yes
Chiasm	D_{max}	<55 Gy	Yes
Optical Nerve	D_{max}	<55 Gy	Yes
Structure	Metric goal in nominal		
Parotid glands	D_{mean}	<26 Gy	Yes
Cochlea	D_{mean}	<45 Gy	Yes
Oral cavity	D_{mean}	<40 Gy	No
Larynx	D_{mean}	<40 Gy	No

Proton planning information: MFO planning; spot spacing size: 5 mm; energy range:

70–250 MeV; range shifter: 5 cm; dose calculation algorithm: pencil beam algorithm (PBA); optimization algorithm: nonlinear universal proton optimizer.

Eclipse version 16.1.0 (Varian Medical Systems, Palo Alto, CA). A relative biological effectiveness (RBE) of 1.1 for proton beams was used. The dosimetric goals for all plans in this study are summarized in the lower part of Table 1. A plan was deemed robust (stop optimization) if the goals set for the clinical target volumes (CTVs) and serial organs are fulfilled for all 12 dose distributions (3 mm orthogonal shifts combined with the $\pm 3.5\%$ range error) as well as the nominal scenario.

2.2 | Extracting small non-rigid variations from CT images

Anatomical variations during the first week of treatment are predominately due to sNRVs, whereas progressive changes (weight-loss, tumor shrinkage) are less significant.^{27–29} Thus, the anatomical changes in the first week from a cohort of patients can be seen representative of a distribution of possible sNRVs.

The sNRVs of a cohort of patients (see Section 2.1) were captured using DIR. DIR finds the optimal deformation vector field (DVF) ϕ to achieve the greatest similarity between two images. We used stationary velocity fields (SVFs) \mathbf{v} of diffeomorphic image registration to identify

anatomical changes in this project. SVFs can easily be calculated from the inverse DVFs ϕ using.³⁰

$$\phi = \exp(\mathbf{v}) \Rightarrow \phi^{-1}(x) = \exp(-\mathbf{v}). \quad (1)$$

To apply the deformations between groups of subjects, we need to project the SVFs into the atlas space, in which all the SVFs have the same position and resolution. The atlas was obtained from a group-wise registration which spatially normalized a cohort of patients^{[1]1,16,31}. In the procedure of the projection, the planning CT (pCT) of each patient was the reference geometry, and the CT acquired during the first treatment week (CT_t) was registered to the pCT to produce $\mathbf{v}_{p \rightarrow t}$, where p stands for pCT and t stands for the week (in this case $t = 1$) when the weekly CT acquired. Then, each patient's pCT was registered to the atlas to produce $\mathbf{v}_{a \rightarrow p}$, where a stands for atlas. $\mathbf{v}_{a \rightarrow p}$ transformed the inter-patient velocity fields $\mathbf{v}_{p \rightarrow t}$ into the atlas using

$$\mathbf{v}_{a,p \rightarrow t} = \mathbf{v}_{a \rightarrow p}^{-1} \circ \mathbf{v}_{p \rightarrow t} \circ \mathbf{v}_{a \rightarrow p}, \quad p \forall P. \quad (2)$$

P includes all the patients' data used in this study.

¹ <https://cmiclab.cs.ucl.ac.uk/mmodat/niftyreg>

Then $\mathbf{v}_{a,p \rightarrow t}$ was transformed into the space of an individual patient \bar{p} using

$$\mathbf{v}_{\bar{p} \rightarrow t} \approx \mathbf{v}_{a \rightarrow \bar{p}}^{-1} \circ \mathbf{v}_{a,p \rightarrow t} \circ \mathbf{v}_{a \rightarrow \bar{p}} \cdot p \forall P. \quad (3)$$

The deformation $\mathbf{v}_{\bar{p} \rightarrow t}$ was used for warping pCT to simulate an sNRV. Finally, in order to warp the planning image $I_{\bar{p}}$, the transformation must be directed from the predicted anatomy to the pCT. This can be simply achieved by reversing the SVFs using

$$\mathbf{v}_{t \rightarrow \bar{p}} = -\mathbf{v}_{\bar{p} \rightarrow t}. \quad (4)$$

The warped image CT^{sNRV} was acquired from:

$$\phi_{t \rightarrow \bar{p}} = \exp(\mathbf{v}_{t \rightarrow \bar{p}}), \quad (5)$$

$$\text{CT}_{\bar{p}}^{\text{sNRV}} = \phi_{t \rightarrow \bar{p}}(\text{pCT}), \quad (6)$$

with $t = 1$ for all the equations above. This method produced 35 CT^{sNRV} s for each patient to represent the possible sNRVs.

The diffeomorphic image registration is implemented in NiftyReg^[1]. NiftyReg is an open-source DIR tool available as part of the NifTK project.

2.3 | Robustness evaluation

We included the 35 sNRV scenarios of each test patient into the robustness evaluation using CT^{sNRV} s. For the four test patients, the dose distributions of IMPT plans were calculated under each robustness scenario. We compared (1) the robust evaluation based on the sNRV scenarios and rigid translation with (2) the conventional setup setting that only includes rigid translation. Probability analysis was used in these two before-treatment evaluations to rank the robustness of IMPT plans for each robustly optimized dose metric listed in the lower part of Table 1.

2.3.1 | Robustness evaluation scenarios

For our proposed evaluation method using the sNRV scenarios (1), we simulated the isocenter shift for each cardinal direction (x_n, y_n, z_n) following the Gaussian distribution with mean $\mu = 0$ mm and standard deviation $\sigma = 1.5$ mm²⁵ on the 35 CT^{sNRV} s. This was done to calculate the perturbed dose distributions caused by the sNRVs and rigid setup uncertainty, since the CT^{sNRV} s have the same isocenter as the planning CT. The 35 dose distributions for each IMPT plan were included in this sNRV+R evaluation.

The conventional evaluation (2) only include the rigid setup uncertainty by applying the same isocenter shifts

used in the sNRV+R evaluation to the planning CT. This way, we achieved 35 perturbed dose distributions per IMPT plan, which were included to evaluate the plan robustness.

2.3.2 | Probability analysis for robust evaluation

The workflow for the sNRV+R evaluation (1) and the conventional evaluation (2) is illustrated in Appendix C. Each considered dose metric D_x (e.g., D_{95}) would have corresponding perturbed dose metrics under the different uncertainty scenarios. The nominal dose metric is subtracted from the perturbed dose metrics to form a distribution of dose metric discrepancies ΔD_x experienced across the uncertainty scenarios.

The upper and lower boundaries of dose metrics in the evaluation can be demonstrated by the shaded areas in the nominal dose–volume histogram (DVH), as an indicator of worst-case scenarios. It was also suggested in the literature to include a probability approach in robust analysis.³² For this, the distance between the probability distribution of ΔD_x under uncertainty and its ideal probability distribution (Dirac delta function, the dose metrics do not change even under uncertainty) was calculated using the Wasserstein distance (WD)

$$\text{WD}(U, I) = \int_{-\infty}^{\infty} |U(x) - I(x)| dx, \quad (7)$$

where U and I are the probability distribution functions of ΔD_x under uncertainty and its ideal distribution, respectively. The WD measures the effort required to convert one distribution into the other. The smaller the WD, the more robust is a plan for this dose metric.

2.3.3 | Performance analysis of robust evaluations

To investigate the effectiveness of sNRVs in indicating the plan robustness to inter-fractional anatomical changes before treatment, the dose discrepancy between accumulated dose using weekly CTs and the nominal dose was taken as the gold standard. In the gold standard evaluation, the dose distributions of the IMPT plans with different beam arrangements were calculated on six weekly CTs of each test patient. Because the accumulated dose is generally used in treatment evaluation and related to prognostics, the weekly dose was accumulated in the reference frame of the planning CT using the DIR algorithm of Niftyreg, referred to as Accu_{Nom} . In the weekly dose calculation, although the isocenter was determined using the information from the rigid registration, the setup error (both rigid setup and sNRV) still existed. Thus, the difference

between Accu_{Nom} and the nominal plan, referred to as ΔD_{st} , represents the influences from total random errors, including actual progression uncertainty and setup uncertainty (both rigid setup and sNRV).

Because different beam arrangements are used in this study, the robustness of beam arrangements can be ranked, referred to as robustness ranking. In the sNRV+R evaluation and the conventional evaluation, the WD is used in robustness ranking for each dose metric. In the gold standard evaluation, ΔD_{st} is used in robustness ranking. To quantitatively validate that the performance of the sNRV+R evaluation is better than the conventional evaluation, the consistency C of the robustness ranking for a dose metric is calculated for each beam arrangement of each patient, using

$$C = |RP_s(B_i) - RP_G(B_i)| - |RP_c(B_i) - RP_G(B_i)|, \quad (8)$$

$RP_s(B_i)$, $RP_c(B_i)$, and $RP_G(B_i)$ represent the robustness ranking position of a beam arrangement B_i in sNRV+R evaluation, conventional evaluation and the gold standard evaluation, respectively. If $C \leq 0$ for a dose metric, then this dose metric of beam arrangement B_i supports that sNRV+R evaluation is better for robust evaluation, compared to the conventional rigid setup evaluation.

3 | RESULT

3.1 | Dosimetric influences caused by small non-rigid variations

This section demonstrates the additional dosimetric influence caused by non-rigid setup uncertainty.

An example of the dose distribution difference caused by an sNRV is shown in Figure 1. The red arrows indicate areas where the dose has fallen under 95% of the prescription dose. With these simulated images of sNRVs and corresponding dose distributions, we can help clinicians to avoid non-rigid postures that can lead to unacceptable dosimetry.

The comparison between the sNRV+R evaluation and the conventional evaluation on an exemplary patient (patient 1) is shown in Figure 2. The upper and lower boundaries of dose metrics in the sNRV+R evaluation (2a) and the conventional evaluation (2b) are indicated by the shaded areas in Figure 2(a,b) separately. We observed that the additional sNRVs widen the bandwidth compared to the conventional robust evaluation. The detailed numbers of the sNRV+R evaluation and the conventional evaluation for four test patients are listed in Appendix D. Among four patients (490 scenarios), we observed a maximum difference in the sNRV+R evaluation to the nominal dose of: 9.37% dose degradation on the D_{95} of CTVs, increase in parotid D_{mean} by 11.87 Gy, increase in larynx D_{mean} by 15.04 Gy,

increase in brainstem D_{max} by 20.82 Gy, increase in spinal cord D_{max} by 20.96 Gy. For CTVs, 4 patients all had scenarios where the CTV D_{95} fell below 95%, 47 out of 490 scenarios in total. In contrast, in conventional evaluation, we observed a maximum difference to the nominal dose of: 7.58% dose degradation on D_{95} of the CTVs, increase in parotid D_{mean} by 4.88 Gy, increase in larynx D_{mean} by 6.13 Gy, increase in brainstem D_{max} by 13.5 Gy, and increase in spinal cord D_{max} by 12.9 Gy.

Please note that the worst-case CTV coverage (D_{95}) under setup uncertainty can drop below 95% in some cases. To generate 35 scenarios in conventional robust evaluation, we let the isocenter shifts follow a Gaussian distribution with mean $\mu = 0$ mm and standard deviation $\sigma = 1.5$ mm. This results in multiple scenarios that can be used for statistical analysis, rather than only using the 12 scenarios usually encountered during robust optimization with 3 mm orthogonal shifts and $\pm 3.5\%$ range error. While we still used the usual 3 mm option to optimize the plan, the additional shifts created with the Gaussian distribution were used for the evaluation. Using this Gaussian distribution may result in scenarios where the shift exceeds 3 mm. However, only 4/490 scenarios were below 95%. Those scenarios only happened to patient 3 whose target volume was located close to the skin, making this particular patient more sensitive to setup uncertainties.

The comparisons of the dose metrics for this patient based on box plots are shown in Figure 2(c)–(h). Dose metrics for the different plans with different beam arrangements are shown in the same figures as box plots. By comparing boxplot of (c)–(e) (sNRV+R evaluation) to (f)–(h) (conventional evaluation) in Figure 2, the mean values of the CTVs' D_{95} in the sNRV+R evaluation are lower than the values in the conventional evaluation ranging from -1.57% to -0.95% (range shows the differences between different beam arrangements). The mean values of parotid D_{mean} , oral cavity D_{mean} , and larynx D_{mean} are higher than the values in conventional evaluation, ranging from 1.02 to 1.82 Gy, 0.52 to 0.70 Gy, and 0.84 to 3.18 Gy, respectively. The mean values of D_{max} of spinal cord, optical nerve, and chiasm between the two evaluations only have slight differences, less than 0.6 Gy.

Figure 2 only partially demonstrates the Dx under uncertainty. In Figure 3, we plot the probability distribution of ΔD_x in the conventional evaluation and in the sNRV+R evaluation on high-risk CTV D_{95} and parotid D_{mean} , respectively, for patient 1. In Figure 3, we can see the influence caused by the sNRVs on the probability distribution of ΔD_x from different beam arrangements. The robustness of a beam arrangement is presented by the closeness of the probability curve of beam arrangements to the Dirac delta function. For the high-risk CTV, the $3B_{60}$ plan is the most robust (the ΔD_{mean} curve of the $3B_{60}$ is the closest

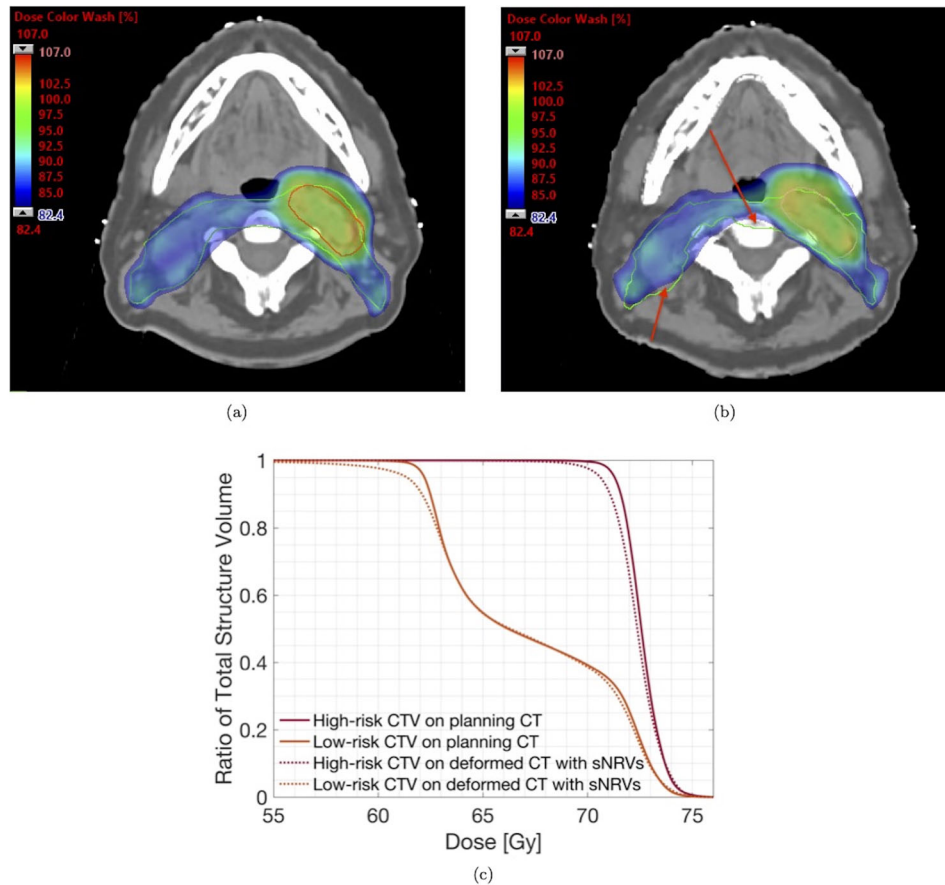


FIGURE 1 An example of dose distribution variations caused by a random small non-rigid variation (sNRV). (a) The dose distribution on the planning CT, and (b) the dose distribution of the same slice on a deformed planning CT. The green contour in images is the low-risk CTV. The color bar was chosen to mask out doses lower than the 95% prescription dose of low-risk CTV (82.4% is corresponding to 95% prescription dose of low-risk CTV). The red arrows indicate areas of underdosage caused by the sNRVs. (c) presents the difference in dose–volume histogram (DVH) caused by the sNRV

to the Dirac delta function, indicated as the dashed vertical line) in the sNRV+R evaluation, as opposed to the conventional evaluation, where we find this beam arrangement to be the less robust one. For the parotid glands, in conventional evaluation (Figure 3c), the $4B_{120}$ is the most robust beam arrangement, while in sNRV+R evaluation (Figure 3d), the $3B_{60}$ is the most robust. The ΔD_{st} from the gold standard evaluation validated that $3B_{60}$ indeed is the most robust beam arrangement for the parotid D_{mean} (please refer to the table in Appendix D).

3.2 | Robust evaluation analysis

This section applies the ranking consistency to demonstrate the benefit of sNRVs for robust evaluation.

Regarding the 10 robustly optimized dose metrics listed in the lower part of Table 1 for each beam arrangement, we calculated the percentage of dose metrics that supports the inclusion of sNRV in robust evaluation based on the data of four test patients. Referring

to the table in Appendix D, we summarized the percentage of dose metrics satisfying $C \leq 0$ ($P_{C \leq 0}$) for each beam arrangement in Table 2. $P_{C \leq 0}$ s are all above 70%, showing that overall including the sNRVs is beneficial to the robust evaluation of all beam arrangements.

We summarized the percentage of $C \leq 0$ across all patients and beam arrangements for each dose metric in Table 3. Here, we can conclude that overall including the sNRVs is beneficial to robust evaluation for each dose metric, compared to only include rigid setup uncertainty.

TABLE 2 Consistency of robustness ranking between two robust evaluation methods for each beam arrangement

Beam arrangements	3B45	3B60	4B110	4B120	5B
$P_{C \leq 0}$ (%)	90	77.5	70	75	72.5

Note: The $P_{C \leq 0}$ summarizes the percentage of the ROI metrics that supports the sNRV+R evaluation ($C \leq 0$) for each beam arrangement.

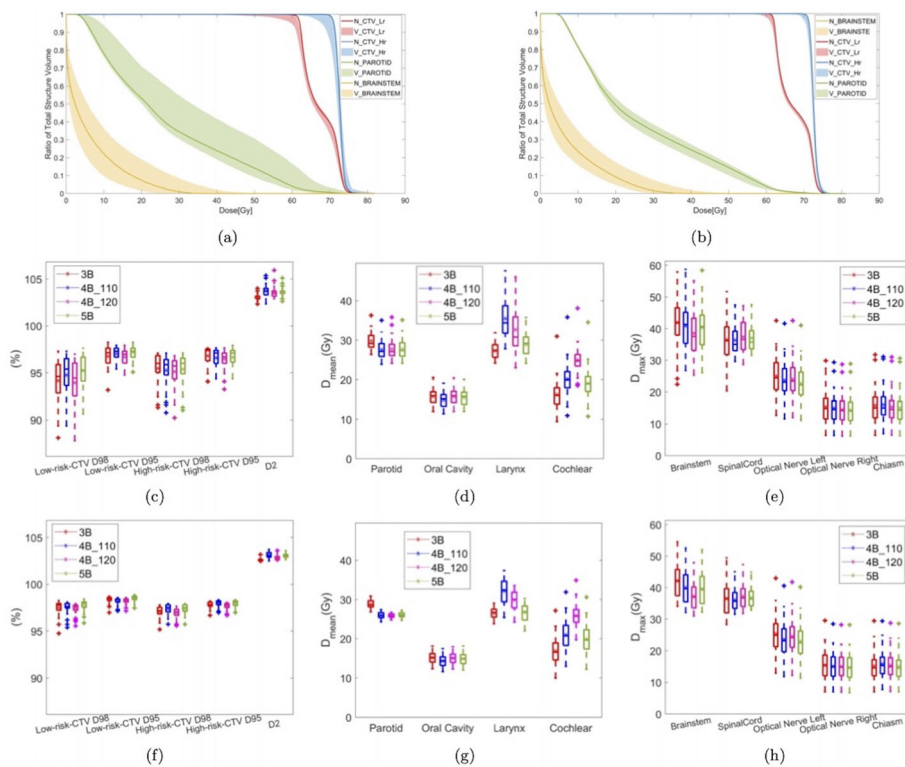


FIGURE 2 The comparison between the sNRV+R evaluation and the conventional evaluation on patient 1. (a,b) The shaded DVH from the 4B₁₂₀ beam arrangement in the sNRV+R evaluation and in the conventional evaluation, respectively. The solid line represents the DVH of the nominal plan (N in the legend), the shaded area indicates the lower and upper boundary of dose metrics in the respective evaluation, caused by the variations (V in the legend). (c–h) Visually summarize the statistics under the respective uncertainty using box plots. The horizontal lines in the box plot indicate the median dose metric among 36 scenarios (including the nominal scenario and 35 uncertainty scenarios defined in robust evaluation). The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. We use the asterisks to indicate the mean value of the dose metrics. (c–e) The boxplots of D_x in the sNRV+R evaluation. (f–h) are the boxplots of D_x in the conventional evaluation

TABLE 3 Consistency of robustness ranking between two robust evaluation methods for each robustly optimized dose metric

Dose metric	High-risk CTV	Low-risk CTV	Parotid	Cochlea	Brainstem	Spinal cord	Chiasm	Optical nerve
	D_{95} (%)	D_{95} (%)	D_{mean} (Gy)	D_{mean} (Gy)	D_{max} (Gy)	D_{max} (Gy)	D_{max} (Gy)	D_{max} (Gy)
$P_{C \leq 0}$ (%)	64.29	78.57	85.71	85.71	64.29	92.86	78.57	71.43

Note: The $P_{C \leq 0}$ summarizes the percentage of $C \leq 0$ over all patients and beam arrangements for each dose metric.

4 | DISCUSSION

Dose distributions in proton therapy are more sensitive to geometric changes than photon therapy. However, in previously published methods of robust evaluation, the impact of anatomical changes before treatment was not considered. In this paper, we demonstrated that including sNRVs into robust evaluation is beneficial.

4.1 | The use of small non-rigid variations for robust beam selection

In the validation of sNRVs' role in robust evaluation, the dose discrepancy that represents the influence from

inter-fractional anatomical changes and isocenter shifts was used as the gold standard. The consistency of robustness ranking showed that $P_{C \leq 0}$ is higher especially on the spinal cord, parotid gland and low-risk CTV, which are closely related to outline changes and neck motions, and also on small structures really sensitive to the sNRVs such as cochlea and chiasm, supporting that sNRVs play a positive role in robust evaluation in terms of indicating robustness to inter-fractional anatomical changes.

The method proposed in this study can assist in selecting robust beam arrangements for proton plans without 4D optimization. In Appendix D, the p -values between the distributions of ΔD_x in the sNRV+R evaluation and conventional evaluation showed that the

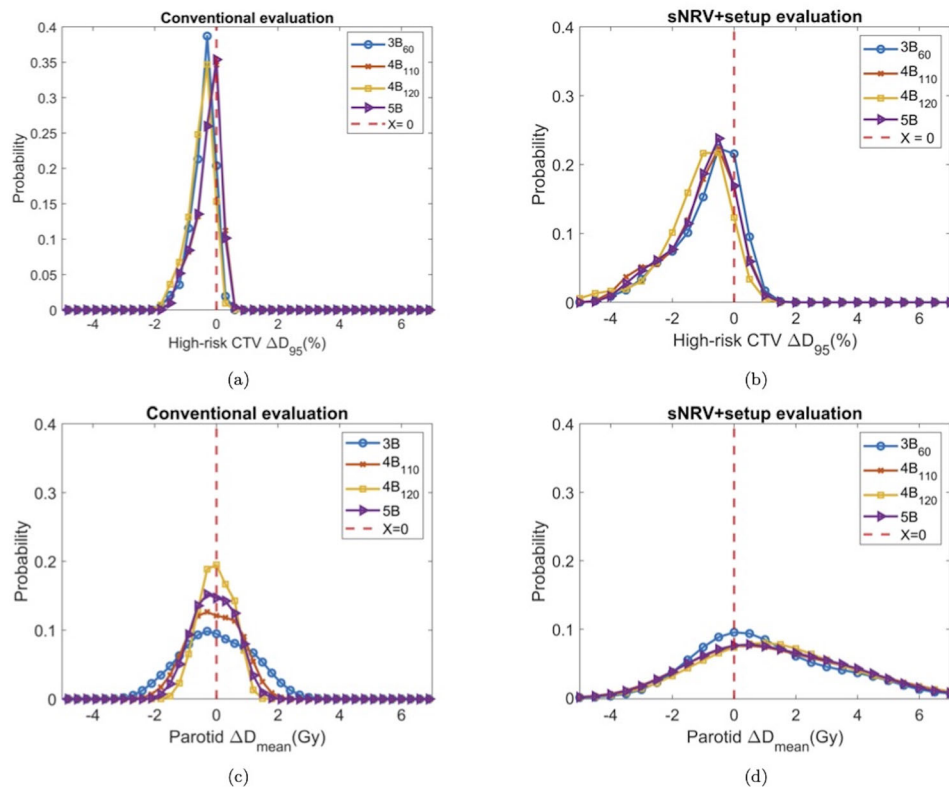


FIGURE 3 Probability distributions of ΔD_x in the conventional evaluation and in the sNRV+R evaluation on patient 1. (a) Probability distribution of ΔD_x in the conventional evaluation on the high-risk CTV D_{95} . (b) Probability distribution of ΔD_x in the sNRV+R evaluation on the high-risk CTV D_{95} . (c) Probability distribution of ΔD_x in the conventional evaluation on the parotid D_{mean} . (d) Probability distribution of ΔD_x in the sNRV+R evaluation on the parotid D_{mean} .

sNRVs mainly influenced the probability distribution of CTVs ΔD_{95} and parotid ΔD_{mean} . The highest priority of the robust optimization for the four test patients in this study was to ensure target coverage. Similar performance of $D_{95\%}$ based on ΔD_{st} was found on different beam arrangements, with differences smaller than 2%. To best demonstrate the advantage of the sNRV+R evaluation over the conventional evaluation, the beam arrangement was selected based on the impact of the sNRVs on the dose of parotid glands as an illustration. Also, the dose on parotid glands is closely related to toxicity such as xerostomia and dysphagia that can have a long-term impact on patients' quality of life. Here, for example, for patient 1, a similar parotid D_{mean} was achieved using $4B_{110}$ and $4B_{120}$. If $4B_{120}$ was selected based on WD, the accumulated parotid D_{mean} reduces by 0.7 Gy, which is corresponding to 1 fraction of D_{mean} delivered to the parotid glands. Other organs can be used for beam selection as well, for example, for patient 1, the rank of the chiasm D_{max} in the sNRV+R evaluation indicated the most robust beam arrangement as the gold standard evaluation.

There were two interesting scenarios worth noticing. In different beam arrangements for patient 1, even

though the nominal parotid D_{mean} of $3B_{60}$ was the highest, the accumulated dose was lower than $4B_{110}$ and $5B$ because $3B_{60}$ was the most robust beam arrangement (the lowest WD) under sNRV+R uncertainty. The ΔD_{st} of $3B_{60}$ showed that $3B_{60}$ controls ΔD_{st} of the parotid D_{mean} within 3 Gy, which is corresponding to 10% NTCP difference³³ and used to trigger replan to protect the parotid glands. This case indicates that beam selection based on robust evaluation can potentially reduce the replan rate, something that needs further investigation in the future. For patient 3, even though the nominal parotid D_{mean} of $4B_{120}$ was higher than in the $5B$ beam arrangements, the accumulated dose was the lowest because $4B_{120}$ was the more robust beam arrangement. A message clearly emerged here is that the best nominal plan may not be the best plan during the treatment.

The impact of different beam angles on the robustness of a plan can be analyzed on the patient-specific geometry using our method. The results can be used to create a robustness plan database to assist to find a more robust planning approach as presented by McGowan et al.²² and Malyapa et al.²³

4.2 | The potential use of small non-rigid variations in clinic

The distribution of sNRVs has the potential to be used in other clinical applications.

First, we found that the sNRV that leads to the most dose discrepancy varies from patient to patient and from beam arrangement to beam arrangement. This approach can help clinicians avoid the set-ups with the sNRVs that can lead to unacceptable dose distributions for a specific patient using a specific beam arrangement.

Second, the acquired sNRVs can potentially assist in better estimating the truly delivered accumulated dose using weekly CTs. The sNRVs can be randomly allocated to each weekly CT, with 5 sNRVs per weekly CT. These deformed weekly CTs can be used to estimate the daily dose distribution under the influence of sNRV. Repeating this procedure can reveal the range of potential accumulated doses for the whole treatment.

Third, this study presented the possibility of including sNRVs from a patient population to robust analysis, which also indicated the potential to be used in robust optimization. Mesías et al.¹⁵ included the first two weekly CTs of patients into robust optimization to account for the sNRVs, suggesting that sNRVs can reduce the need for adaption. They indicated that the first two weekly CTs can be replaced by a series of CT images scanned before treatment. Li et al.³⁴ considered weekly CTs in the robust evaluation, and Yang et al.³⁵ added the adaptive planning CTs into robust optimization. However, their methods relied on the acquisition of CT images during the treatment, which limits the creation of a robust plan at the early planning stage. In contrast to their patient-specific approach, I suggest an atlas-based technique. While this approach is not patient-specific but based on the assumption that sNRVs are mainly random, there are some advantages. First, this method does not require the acquisition of a series of CT images of the same patient pre-treatment, therefore saving imaging dose and reducing workload. Second, assuming that sNRVs can be reasonably represented using this method, deformed images with the sNRVs can be prepared in advance and fully exploit the benefits of robust optimization with multiple CTs. This will be investigated in future studies.

It should be mentioned that the inclusion of a large patient cohort (many sNRV scenarios) would require recalculating the treatment plan many times. For efficiency, I suggest limiting the number of included sNRVs to the most common/frequent ones. The most common sNRVs can be found, for example, by using anatomical models which use principal component analysis applied to anatomical deformations of a patient cohort to estimate the likelihood of a certain anatomical deformation to happen. By only including the most likely principal components of the deformation into the robust evalu-

ation, the number of recalculated plans can be reduced while still representing well the sNRVs. This trade-off will be explored in future work.

The concept presented here can be adapted to different scenarios. Here, I did not factor in immobilization equipment and patient characteristics such as age, size, disease staging, and physical condition. All those factors are likely to influence the possibility and the amplitude of a specific anatomical change to arise during the treatment. While this is not yet considered in this paper, the presented approach has the potential to do so. If sufficient patient data are used to build the atlas, the patient data can be stratified into groups of different immobilization devices and patient characteristics before performing the robust evaluation.

4.3 | Limitations

For the purpose of showing the feasibility, different plans with different beam arrangements were only created for four test patients. Further validation of the methods will be conducted on a large number of patients. Another limitation of this work is that the impact of DIR accuracy was ignored on robust evaluation. We assumed that the DIR uncertainty would equally affect the robust evaluation for different beam arrangements.

When validating the role of sNRVs in robust evaluation, we decided to not take the range uncertainty from Hounsfield units (HU) into account because range uncertainty is an isolated source considered in the robust evaluation and is solely based on the CT calibration. Therefore, it should only have small influence on the results of the comparison, which established that sNRV should be considered in the robust evaluation as a component of the random set up errors, not just rigid set up. However, to fully evaluate the plan, the range uncertainty should be used with the translation rigid setup and sNRVs. Please see [Appendix E](#) for an example of the further evaluation.

5 | CONCLUSION

This study demonstrated the additional dose discrepancy arising from sNRVs and the robust plan evaluation method based on the probability distribution. The novelty of this study exists in three aspects: (1) compared with the conventional evaluation, we demonstrated that the inclusion of sNRVs can be beneficial to robust evaluation in terms of indicating plan robustness to inter-fractional anatomical changes. (2) This atlas-based method can help clinicians to choose plans that are robust against those sNRVs. (3) Our method can potentially provide multiple images for potential future 4D optimization.

ACKNOWLEDGMENTS

Ying Zhang is supported by a scholarship from the China Scholarship Council Nos. 201809150003. Esther Bär is supported by the Radiation Research Unit at the Cancer Research UK City of London Centre Award C7893/A28990. Richard A. Amos is supported by a Cancer Research UK Centres Network Accelerator Award Grant (A21993) to the 3 ART-NET Consortium. Wenyong Tan is supported by the National Natural Science Funding of China (81974462).

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

Restricted access. The authors cannot make these data publicly available due to data use agreement.

REFERENCES

- Steneker M, Lomax A, Schneider U. Intensity modulated photon and proton therapy for the treatment of head and neck tumors. *Radiother Oncol.* 2006;80(2):263-267.
- Van De Water TA, Lomax AJ, Bijl HP, et al. Potential benefits of scanned intensity-modulated proton therapy versus advanced photon therapy with regard to sparing of the salivary glands in oropharyngeal cancer. *Int J Radiat Oncol Biol Phys.* 2011;79(4):1216-1224.
- McGowan SE, Burnet NG, Lomax AJ. Treatment planning optimisation in proton therapy. *Br J Radiol.* 2013;86:20120288-20120288.
- Minatogawa H, Yasuda K, Dekura Y, et al. Potential benefits of adaptive intensity-modulated proton therapy in nasopharyngeal carcinomas. *J Appl Clin Med Phys.* 2021;22(1):174-183.
- Lester-Coll NH, Margalit DN. Modeling the potential benefits of proton therapy for patients with oropharyngeal head and neck cancer. *Int J Radiat Oncol Biol Phys.* 2019;104(3):563-566.
- Water TA, Bijl HP, Schilstra C, Pijls-Johannesma M, Langendijk JA. The potential benefit of radiotherapy with protons in head and neck cancer with respect to normal tissue sparing: A systematic review of literature. *Oncologist.* 2011;16(3).
- Wu RY, Liu AY, Sio TT, et al. Intensity-modulated proton therapy adaptive planning for patients with oropharyngeal cancer. *Int J Part Ther.* 2017;4(2):26-34.
- Kraan AC, de Water S, Teguh DN, et al. Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors. *Int J Radiat Oncol Biol Phys.* 2013;87(5):888.
- Yan DF, Yan SX, Wang QD, Liao XB, Lu ZJ, Wang YX. Predictors for replanning in loco-regionally advanced nasopharyngeal carcinoma patients undergoing intensity-modulated radiation therapy: A prospective observational study. *BMC Cancer.* 2013; 13(1):1-9.
- Tan W, Li Y, Han G. Target volume and position variations during intensity-modulated radiotherapy for patients with nasopharyngeal carcinoma. *Oncol Targets Ther.* 2013;6:1719-1728.
- Bhide SA, Davies M, Burke K, et al. Weekly volume and dosimetric changes during chemoradiotherapy with intensity-modulated radiation therapy for head and neck cancer: A prospective observational study. *Int J Radiat Oncol Biol Phys.* 2010;76(5):1360-1368.
- Hansen EK, Bucci MK, Quivey JM, Weinberg V, Xia P. Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys.* 2006;64(2):355-362.
- Bruijnen T, Stemkens B, Terhaard CHJ, Lagendijk JJW, Raaijmakers CPJ, Tjissen RHN. Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. *Radiother Oncol.* 2019;130:82-88.
- Brouwer CL, Steenbakkers RJHM, van der Schaaf A, et al. Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. *Radiother Oncol.* 2016;120(1):36-40.
- Cubillos-Mesías M, Troost EGC, Lohaus F, et al. Including anatomical variations in robust optimization for head and neck proton therapy can reduce the need of adaptation. *Radiother Oncol.* 2019;131:127-134.
- Zhang Y, McGowan Holloway S, Wilson MZ, et al. DIR-based models to predict weekly anatomical changes in head and neck cancer proton therapy. *Phys Med Biol.* 2022;67(9).
- Zhang Y, Alshaiqi J, Amos RA, et al. Improving workflow for adaptive proton therapy with predictive anatomical modelling: A proof of concept. *Radiother Oncol.* 2022;173:93-101.
- Matter M, Nenoff L, Meier G, Weber DC, Lomax AJ, Albertini F. Intensity modulated proton therapy plan generation in under ten seconds. *Acta Oncol.* 2019;58(10):1435-1439.
- Yepes P, Adair A, Grosshans D, et al. Comparison of Monte Carlo and analytical dose computations for intensity modulated proton therapy. *Phys Med Biol.* 2018;63(4):045003.
- Bobić M, Lalonde A, Sharp GC, et al. Comparison of weekly and daily online adaptation for head and neck intensity-modulated proton therapy. *Phys Med Biol.* 2021;66:055023.
- Lalonde A, Bobić M, Winey B, Verburg J, Sharp GC, Paganetti H. Anatomic changes in head and neck intensity-modulated proton therapy: Comparison between robust optimization and online adaptation. *Radiother Oncol.* 2021;159:39-47.
- McGowan SE, Albertini F, Thomas SJ, Lomax AJ. Defining robustness protocols: A method to include and evaluate robustness in clinical plans. *Phys Med Biol.* 2015;60(7):2671-2684.
- Malyapa R, Lowe M, Bolsi A, Lomax AJ, Weber DC, Albertini F. Evaluation of robustness to setup and range uncertainties for head and neck patients treated with pencil beam scanning proton therapy. *Int J Radiat Oncol Biol Phys.* 2016;95(1):154-162.
- Hague C, Aznar M, Dong L, et al. Inter-fraction robustness of intensity-modulated proton therapy in the post-operative treatment of oropharyngeal and oral cavity squamous cell carcinomas. *Br J Radiol.* 2020;93(1107):20190638.
- Cubillos-Mesías M, Troost EGC, Lohaus F, et al. Quantification of plan robustness against different uncertainty sources for classical and anatomical robust optimized treatment plans in head and neck cancer proton therapy. *Br J Radiol.* 2020;93(1107):20190573.
- Holloway SM, Holloway MD, Thomas SJ. A method for acquiring random range uncertainty probability distributions in proton therapy. *Phys Med Biol.* 2017;63(1):01NT02.
- Barker JL, Garden AS, Ang KK, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *Int J Radiat Oncol Biol Phys.* 2004;59(4):960-970.
- Lee C, Langen KM, Lu W, et al. Evaluation of geometric changes of parotid glands during head and neck cancer radiotherapy using daily MVCT and automatic deformable registration. *Radiother Oncol.* 2008;89(1):1563-1571.
- Stützer K, Jakobi A, Bandurska-Luque A, et al. Potential proton and photon dose degradation in advanced head and neck cancer patients by intrathrapy changes. *J Appl Clin Med Phys.* 2017;18(6):104-113.
- Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal.* 2008;12(1):26-41.

31. Veiga C, Lim P, Anaya VM, et al. Atlas construction and spatial normalisation to facilitate radiation-induced late effects research in childhood cancer. *Phys Med Biol*. 2021;66(10):105005.
32. Hernandez V, Hansen CR, Widesott L, et al. What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans. *Radiother Oncol*. 2020;153:26-33.
33. Brouwer CL, Steenbakkens RJHM, van der Schaaf A, et al. Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. *Radiother Oncol*. 2016;120(1):36-40.
34. Li H, Zhang X, Park P, et al. Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients. *Radiother Oncol*. 2015;114(3):367-372.
35. Yang Z, Zhang X, Wang X, et al. Multiple-CT optimization: An adaptive optimization method to account for anatomical changes in intensity-modulated proton therapy for head and neck cancers. *Radiother Oncol*. 2020;142:124-132.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhang Y, Alshaikhi J, Amos RA, et al. Pre-treatment analysis of non-rigid variations can assist robust intensity-modulated proton therapy plan selection for head and neck patients. *Med Phys*. 2022;49:7683–7693.
<https://doi.org/10.1002/mp.15971>